

A Bayesian-motivated test for linear model in high-dimensional setting

Rui Wang

Friday 14th December, 2018

1 Introduction

The proposed test is the limit of Bayes factors.

Fixed design

Suppose we would like to test the hypotheses:

$$\begin{aligned}\mathcal{H}_0 : \mathbf{y} &= \mathbf{X}_a \boldsymbol{\beta}_a + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}_n(0, \phi^{-1} \mathbf{I}_n), \\ \mathcal{H}_1 : \mathbf{y} &= \mathbf{X}_a \boldsymbol{\beta}_a + \mathbf{X}_b \boldsymbol{\beta}_b + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}_n(0, \phi^{-1} \mathbf{I}_n).\end{aligned}$$

Here $\boldsymbol{\beta}_a$ is q dimensional and $\boldsymbol{\beta}_b$ is p dimensional. We assume that as n tends to infinity, q is fixed while $p/n \rightarrow \infty$. This assumption is reasonable. We assume \mathbf{X}_a has full column rank and \mathbf{X}_b has full row rank. In practice, p_0 is often 1 and \mathbf{X}_a is $\mathbf{1}_n$.

As Goeman et al. (2006) pointed out, if $\boldsymbol{\beta}_b \neq 0$ but $\mathbf{X}_b \boldsymbol{\beta}_b = 0$, no test has any power. Goeman et al. (2006) used Bayesian method. Their idea is to choose an ‘unbiased’ distribution of $\boldsymbol{\beta}_b$. As they noticed, their test has negligible power for many alternatives, and is not unbiased.

The following proposition implies that there is no nontrivial unbiased test.

Proposition 1. *Suppose $\mathbf{y} \sim \mathcal{N}_n(\mu, \phi^{-1} \mathbf{I}_n)$. We test $H_0 : \mu = \mathbf{X}_a \boldsymbol{\beta}_a, \boldsymbol{\beta}_a \in \mathbb{R}^q$ versus $H_1 : \mu \in \mathbb{R}^n$, where \mathbf{X}_a is an $n \times q$ matrix with full column rank, $q < n$. Let $\varphi(\mathbf{y})$ be a test function, that is, a Borel measurable function, $0 \leq \varphi(\mathbf{y}) \leq 1$. If $\int \varphi(\mathbf{y}) \mathcal{N}_n(\mathbf{X}_a \boldsymbol{\beta}_a, \phi^{-1} \mathbf{I}_n)(d\mathbf{y}) = \alpha$ for $\boldsymbol{\beta}_a \in \mathbb{R}^q$, $\phi > 0$ and $\int \varphi(\mathbf{y}) \mathcal{N}_n(\mu, \phi^{-1} \mathbf{I}_n)(d\mathbf{y}) \geq \alpha$ for $\mu \in \mathbb{R}^n$, $\phi > 0$, then $\varphi(\mathbf{y}) = \alpha$, a.s.*

So we can not find a universally good test. Instead, we would like to find a test with good average behaviour. So Bayesian methods are natural choices in this case.

Bayes hypothesis testing use the Bayes factor.

$$B_{10} = \frac{\int f_1(\mathbf{y} | \boldsymbol{\beta}_b, \boldsymbol{\beta}_a, \phi) \pi_1(\boldsymbol{\beta}_b, \boldsymbol{\beta}_a, \phi) d\boldsymbol{\beta}_b d\boldsymbol{\beta}_a d\phi}{\int f_0(\mathbf{y} | \boldsymbol{\beta}_a, \phi) \pi_0(\boldsymbol{\beta}_a, \phi) d\boldsymbol{\beta}_a d\phi}.$$

There have been several extensions of g -priors to $p > n$ case: Maruyama and George (2011), Shang and Clayton (2011).

Under H_0 , we impose the reference prior $\pi_0(\beta_a, \phi) = c/\phi$. Note that under H_1 , the posterior corresponding to the reference prior is proper if and only if $\text{Rank}(\mathbf{X}_a, \mathbf{X}_b) = q + p$ and $n > q + p$. That is, the minimal training sample size is $q + p + 1$. So we cannot impose the reference prior under H_1 provided $q + p \geq n$. We temporarily impose the conditional prior $\beta_b|\beta_a, \phi \sim \mathcal{N}_p(0, \kappa^{-1}\phi^{-1}\mathbf{I}_p)$. There are extensive literature consider the choice of κ . Kass and Wasserman (1995) choose κ such that the amount of information about the parameter equal to the amount of information contained in one observation. Thus, under H_1 , we put the prior

$$\pi_1(\beta_b|\beta_a, \phi) = \mathcal{N}_p\left(0, \frac{1}{\kappa\phi}\mathbf{I}_p\right)(\beta_b), \quad \pi_1(\beta_a, \phi) = \frac{c}{\phi}.$$

$$\begin{aligned} m_0(\mathbf{y}; \kappa, \tau) &:= \int f_0^\tau(\mathbf{y}|\beta_a, \phi)\pi_0(\beta_a, \phi)d\beta_a d\phi \\ &= \frac{c_0\Gamma\left(\frac{\tau n - q}{2}\right)}{\pi^{\frac{\tau n - q}{2}}\tau^{\frac{\tau n}{2}}|\mathbf{X}_a^\top \mathbf{X}_a|^{\frac{1}{2}}\|(\mathbf{I}_n - \mathbf{P}_a)\mathbf{y}\|^{\tau n - q}}. \end{aligned}$$

$$\begin{aligned} m_1(\mathbf{y}; \kappa, \tau) &:= \int f_1^\tau(\mathbf{y}|\beta_b, \beta_a, \phi)\pi_1(\beta_b|\beta_a, \phi)\pi_1(\beta_a, \phi)d\beta_a d\beta_b d\phi \\ &= \frac{c_1\kappa^{\frac{p}{2}}\Gamma\left(\frac{\tau n - q}{2}\right)}{\pi^{\frac{\tau n - q}{2}}\tau^{\frac{\tau n + p}{2}}|\mathbf{X}_a^\top \mathbf{X}_a|^{\frac{1}{2}}|\mathbf{X}_b^{*\top} \mathbf{X}_b^* + \frac{\kappa}{\tau}\mathbf{I}_p|^{\frac{1}{2}}}\frac{1}{[\mathbf{y}^{*\top} \mathbf{y}^* - \mathbf{y}^{*\top} \mathbf{X}_b^*(\mathbf{X}_b^{*\top} \mathbf{X}_b^* + \frac{\kappa}{\tau}\mathbf{I}_p)^{-1}\mathbf{X}_b^{*\top} \mathbf{y}^*]^{\frac{\tau n - q}{2}}}. \\ \frac{m_1(\mathbf{y}; \kappa, \tau)}{m_0(\mathbf{y}; \kappa, \tau)} &= \frac{c_1\kappa^{\frac{p}{2}}}{c_0\tau^{\frac{p}{2}}|\mathbf{X}_b^{*\top} \mathbf{X}_b^* + \frac{\kappa}{\tau}\mathbf{I}_p|^{\frac{1}{2}}}\left(\frac{\mathbf{y}^{*\top} \mathbf{y}^*}{\mathbf{y}^{*\top} \mathbf{y}^* - \mathbf{y}^{*\top} \mathbf{X}_b^*(\mathbf{X}_b^{*\top} \mathbf{X}_b^* + \frac{\kappa}{\tau}\mathbf{I}_p)^{-1}\mathbf{X}_b^{*\top} \mathbf{y}^*}\right)^{\frac{\tau n - q}{2}} \end{aligned}$$

It is straightforward to show that the Bayes factor associated with these priors is

$$\begin{aligned} B_{10}^\kappa &= \frac{\kappa^{p/2}}{|\mathbf{X}_b^\top (\mathbf{I}_n - \mathbf{P}_a) \mathbf{X}_b + \kappa \mathbf{I}_p|^{1/2}} \\ &\quad \left(\frac{\mathbf{y}^\top (\mathbf{I}_n - \mathbf{P}_a) \mathbf{y}}{\mathbf{y}^\top (\mathbf{I}_n - \mathbf{P}_a) \mathbf{y} - \mathbf{y}^\top (\mathbf{I}_n - \mathbf{P}_a) \mathbf{X}_b (\mathbf{X}_b^\top (\mathbf{I} - \mathbf{P}_a) \mathbf{X}_b + \kappa \mathbf{I}_p)^{-1} \mathbf{X}_b^\top (\mathbf{I}_n - \mathbf{P}_a) \mathbf{y}} \right)^{(n-q)/2}. \end{aligned}$$

Thus,

$$\begin{aligned} 2 \log B_{10}^\kappa &= p \log \kappa - \log |\mathbf{X}_b^\top (\mathbf{I}_n - \mathbf{P}_a) \mathbf{X}_b + \kappa \mathbf{I}_p| \\ &\quad - (n - q) \log \left(1 - \frac{\mathbf{y}^\top (\mathbf{I}_n - \mathbf{P}_a) \mathbf{X}_b (\mathbf{X}_b^\top (\mathbf{I} - \mathbf{P}_a) \mathbf{X}_b + \kappa \mathbf{I}_p)^{-1} \mathbf{X}_b^\top (\mathbf{I}_n - \mathbf{P}_a) \mathbf{y}}{\mathbf{y}^\top (\mathbf{I}_n - \mathbf{P}_a) \mathbf{y}} \right). \end{aligned}$$

Denote by $\mathbf{I}_n - \mathbf{P}_a = \tilde{\mathbf{U}}_a \tilde{\mathbf{U}}_a^\top$ the rank decomposition of $\mathbf{I}_n - \mathbf{P}_a$, where $\tilde{\mathbf{U}}_a$ is a $n \times (n - q)$ column orthogonal matrix. Let $\mathbf{X}_b^* = \tilde{\mathbf{U}}_a^\top \mathbf{X}_b$, $\mathbf{y}^* = \tilde{\mathbf{U}}_a^\top \mathbf{y}$. Let γ_i be the i th largest eigenvalue of $\mathbf{X}_b^* \mathbf{X}_b^{*\top}$, $i = 1, \dots, n - q$. Denote by $\mathbf{X}_b^* = \mathbf{U}_b^* \mathbf{D}_b^* \mathbf{V}_b^{*\top}$ the singular value decomposition of \mathbf{X}_b^* , where \mathbf{U}_b^* , \mathbf{V}_b^* are $(n - q) \times (n - q)$ and $p \times (n - q)$ column orthogonal matrices, respectively, and $\mathbf{D}_b^* = \text{diag}(\sqrt{\gamma_1}, \dots, \sqrt{\gamma_{n-q}})$. Then

$$\begin{aligned}
2 \log B_{10}^\kappa &= p \log \kappa - \sum_{i=1}^{n-q} \log(\gamma_i + \kappa) - (p - (n - q)) \log \kappa \\
&\quad - (n - q) \log \left(1 - \frac{\mathbf{y}^{*\top} \mathbf{X}_b^* (\mathbf{X}_b^{*\top} \mathbf{X}_b^* + \kappa \mathbf{I}_p)^{-1} \mathbf{X}_b^{*\top} \mathbf{y}^*}{\mathbf{y}^{*\top} \mathbf{y}^*} \right) \\
&= - \sum_{i=1}^{n-q} \log(\gamma_i + \kappa) + (n - q) \log \left(\frac{\mathbf{y}^{*\top} \mathbf{y}^*}{\mathbf{y}^{*\top} \mathbf{U}_b^* \left[\frac{1}{\kappa} (\mathbf{I}_{n-q} - \mathbf{D}_b^* (\mathbf{D}_b^{*2} + \kappa \mathbf{I}_{n-q})^{-1} \mathbf{D}_b^*) \right] \mathbf{U}_b^{*\top} \mathbf{y}^*} \right) \\
&= (n - q) \log \kappa - \sum_{i=1}^{n-q} \log(\gamma_i + \kappa) - (n - q) \log \left(1 - \frac{\mathbf{y}^{*\top} \mathbf{U}_b^* \mathbf{D}_b^* (\mathbf{D}_b^{*2} + \kappa \mathbf{I}_{n-q})^{-1} \mathbf{D}_b^* \mathbf{U}_b^{*\top} \mathbf{y}^*}{\mathbf{y}^{*\top} \mathbf{y}^*} \right).
\end{aligned}$$

The main part of $2 \log B_{10}^\kappa$ is

$$T_n^\kappa = \frac{\mathbf{y}^{*\top} \mathbf{U}_b^* \mathbf{D}_b^* (\mathbf{D}_b^{*2} + \kappa \mathbf{I}_{n-q})^{-1} \mathbf{D}_b^* \mathbf{U}_b^{*\top} \mathbf{y}^*}{\mathbf{y}^{*\top} \mathbf{y}^*}.$$

A large value of T_n^κ supports the alternative hypothesis. Under the null hypothesis,

$$\mathbb{E} T_n^\kappa = \frac{1}{n - q} \text{tr} (\mathbf{D}_b^{*2} (\mathbf{D}_b^{*2} + \kappa \mathbf{I}_{n-q})^{-1}).$$

Under the alternative hypothesis, consider $\beta_b = c \beta_b^\dagger$ where $\beta_b^\dagger \neq 0$ is a fixed direction and $c > 0$.

As $c \rightarrow \infty$,

$$T_n^\kappa \rightarrow \frac{\beta_b^{\dagger\top} \mathbf{V}_b^* \mathbf{D}_b^{*2} (\mathbf{D}_b^{*2} + \kappa \mathbf{I}_{n-q})^{-1} \mathbf{D}_b^{*2} \mathbf{V}_b^{*\top} \beta_b^\dagger}{\beta_b^{\dagger\top} \mathbf{V}_b^* \mathbf{D}_b^{*2} \mathbf{V}_b^{*\top} \beta_b^\dagger}.$$

We say T_n^κ is consistent along the direction β_b^\dagger if

$$\frac{\beta_b^{\dagger\top} \mathbf{V}_b^* \mathbf{D}_b^{*2} (\mathbf{D}_b^{*2} + \kappa \mathbf{I}_{n-q})^{-1} \mathbf{D}_b^{*2} \mathbf{V}_b^{*\top} \beta_b^\dagger}{\beta_b^{\dagger\top} \mathbf{V}_b^* \mathbf{D}_b^{*2} \mathbf{V}_b^{*\top} \beta_b^\dagger} > \frac{1}{n - q} \text{tr} (\mathbf{D}_b^{*2} (\mathbf{D}_b^{*2} + \kappa \mathbf{I}_{n-q})^{-1}),$$

or equivalently

$$\beta_b^{\dagger\top} \mathbf{V}_b^* \left[\mathbf{D}_b^{*2} (\mathbf{D}_b^{*2} + \kappa \mathbf{I}_{n-q})^{-1} \mathbf{D}_b^{*2} - \frac{1}{n - q} \text{tr} (\mathbf{D}_b^{*2} (\mathbf{D}_b^{*2} + \kappa \mathbf{I}_{n-q})^{-1}) \mathbf{D}_b^{*2} \right] \mathbf{V}_b^{*\top} \beta_b^\dagger > 0.$$

Let k_κ be the number of positive eigenvalues of

$$\mathbf{V}_b^* \left[\mathbf{D}_b^{*2} (\mathbf{D}_b^{*2} + \kappa \mathbf{I}_{n-q})^{-1} \mathbf{D}_b^{*2} - \frac{1}{n - q} \text{tr} (\mathbf{D}_b^{*2} (\mathbf{D}_b^{*2} + \kappa \mathbf{I}_{n-q})^{-1}) \mathbf{D}_b^{*2} \right] \mathbf{V}_b^{*\top}.$$

Let \mathcal{S}_κ be the linear space spanned by the first k_κ columns of \mathbf{V}_b^* . Denote by \mathcal{S}_κ^\perp the orthogonal complement space of \mathcal{S}_κ . We have $\mathbb{R}^p = \mathcal{S}_\kappa \oplus \mathcal{S}_\kappa^\perp$. If $\beta_b^\dagger \in \mathcal{S}_\kappa$,

$$\mathbf{V}_b^* \left[\mathbf{D}_b^{*2} (\mathbf{D}_b^{*2} + \kappa \mathbf{I}_{n-q})^{-1} \mathbf{D}_b^{*2} - \frac{1}{n - q} \text{tr} (\mathbf{D}_b^{*2} (\mathbf{D}_b^{*2} + \kappa \mathbf{I}_{n-q})^{-1}) \mathbf{D}_b^{*2} \right] \mathbf{V}_b^{*\top} > 0.$$

On the other hand, if $\beta_b^\dagger \in \mathcal{S}_\kappa^\perp$,

$$\mathbf{V}_b^* \left[\mathbf{D}_b^{*2} (\mathbf{D}_b^{*2} + \kappa \mathbf{I}_{n-q})^{-1} \mathbf{D}_b^{*2} - \frac{1}{n-q} \text{tr} (\mathbf{D}_b^{*2} (\mathbf{D}_b^{*2} + \kappa \mathbf{I}_{n-q})^{-1}) \mathbf{D}_b^{*2} \right] \mathbf{V}_b^{*\top} \leq 0.$$

We would like to choose a hyperparameter κ which consists the most consistent directions. To achieve this, we maximize k_κ with respect to κ .

Proposition 2. *For $\kappa_2 > \kappa_1 > 0$, we have $k_{\kappa_1} \geq k_{\kappa_2}$. That is, k_κ ($\kappa > 0$) is decreasing in κ .*

The proposition implies that we should put κ as small as possible. This motivates us to consider $B_{10}^0 = \lim_{\kappa \rightarrow 0} B_{10}^\kappa$. It is straightforward to show that

$$2 \log B_{10}^0 = - \sum_{i=1}^{n-q} \log(\gamma_i) + (n-q) \log \left(\frac{\mathbf{y}^{*\top} \mathbf{y}^*}{\mathbf{y}^{*\top} (\mathbf{X}_b^* \mathbf{X}_b^{*\top})^{-1} \mathbf{y}^*} \right).$$

B_{10}^0 can be regarded as the Bayes factor with respect to noninformative prior.

Define

$$T_n = \frac{\mathbf{y}^{*\top} (\mathbf{X}_b^* \mathbf{X}_b^{*\top})^{-1} \mathbf{y}^*}{\mathbf{y}^{*\top} \mathbf{y}^*}.$$

Then we reject the null hypothesis if T_n is small. It can be seen that under the null hypothesis,

$$T_n \sim \frac{\sum_{i=1}^{n-q} \gamma_i^{-1} Z_i^2}{\sum_{i=1}^{n-q} Z_i^2},$$

where γ_i is the i th eigenvalue of $\mathbf{X}_b^* \mathbf{X}_b^{*\top}$, $i = 1, \dots, n-q$, and Z_1, \dots, Z_{n-q} are iid $\mathcal{N}(0, 1)$ random variables.

2 Asymptotic results

Let $\boldsymbol{\varepsilon} = (\epsilon_1, \dots, \epsilon_n)^\top$, where ϵ_i 's are iid random variable. Denote $\mu_k = \mathbb{E} \epsilon_1^k$. Then $\mu_1 = 0$, $\mu_2 = \phi^{-1}$.

Assumption 1. *Suppose*

Lemma 1. *If $\phi^2 \mu_4 = o(n-q)$,*

$$\mathbf{y}^{*\top} \mathbf{y}^* = (1 + o_P(1)) \left(\beta_b^\top \mathbf{X}_b^\top (\mathbf{I}_n - \mathbf{P}_a) \mathbf{X}_b \beta_b + \phi^{-1} (n-q) \right).$$

Proof.

$$\mathbf{y}^{*\top} \mathbf{y}^* = \beta_b^\top \mathbf{X}_b^\top (\mathbf{I}_n - \mathbf{P}_a) \mathbf{X}_b \beta_b + 2\boldsymbol{\varepsilon}^\top (\mathbf{I}_n - \mathbf{P}_a) \mathbf{X}_b \beta_b + \boldsymbol{\varepsilon}^\top (\mathbf{I}_n - \mathbf{P}_a) \boldsymbol{\varepsilon}.$$

$$\mathbb{E} \left(\mathbf{y}^{*\top} \mathbf{y}^* \right) = \beta_b^\top \mathbf{X}_b^\top (\mathbf{I}_n - \mathbf{P}_a) \mathbf{X}_b \beta_b + \phi^{-1} (n-q).$$

$$\text{Var} \left(\mathbf{y}^{*\top} \mathbf{y}^* \right) \leq 2 \text{Var} \left(2\boldsymbol{\varepsilon}^\top (\mathbf{I}_n - \mathbf{P}_a) \mathbf{X}_b \beta_b \right) + 2 \text{Var} \left(\boldsymbol{\varepsilon}^\top (\mathbf{I}_n - \mathbf{P}_a) \boldsymbol{\varepsilon} \right)$$

From (i) of (Chen et al., 2010, Proposition A.1),

$$\text{Var}\left(\varepsilon^\top (\mathbf{I}_n - \mathbf{P}_a) \varepsilon\right) = \phi^{-2} \left((\phi^2 \mu_4 - 3) \sum_{i=1}^n ((\mathbf{I}_n - \mathbf{P}_a)_{i,i})^2 + 2(n-q) \right) \leq \phi^{-2} (2 + \phi^2 \mu_4) (n-q).$$

Then

$$\text{Var}\left(\mathbf{y}^{*\top} \mathbf{y}^*\right) \leq 8\phi^{-1} \boldsymbol{\beta}_b^\top \mathbf{X}_b^\top (\mathbf{I}_n - \mathbf{P}_a) \mathbf{X}_b \boldsymbol{\beta}_b + 2\phi^{-2} (2 + \phi^2 \mu_4) (n-q)$$

Thus, if $\phi^2 \mu_4 = o(n-q)$, we have

$$\frac{\text{Var}\left(\mathbf{y}^{*\top} \mathbf{y}^*\right)}{(\mathbb{E}(\mathbf{y}^{*\top} \mathbf{y}^*))^2} \rightarrow 0,$$

and consequently $\mathbf{y}^{*\top} \mathbf{y}^* = (1 + o_P(1)) \mathbb{E}(\mathbf{y}^{*\top} \mathbf{y}^*)$.

□

Note that under the normality, $T_n - \text{tr}((\mathbf{X}_b^* \mathbf{X}_b^{*\top})^{-1})/(n-q)$ has zero mean.

Theorem 1. *Let \mathbf{A}_n be an $(n-q) \times (n-q)$ symmetric matrix.*

$$\left(\boldsymbol{\beta}_b^\top \mathbf{X}_b^\top (\mathbf{I}_n - \mathbf{P}_a) \mathbf{X}_b \boldsymbol{\beta}_b + \phi^{-1} (n-q) \right) \left(\frac{\mathbf{y}^{*\top} \mathbf{A}_n \mathbf{y}^*}{\mathbf{y}^{*\top} \mathbf{y}^*} - \frac{\text{tr}(\mathbf{A}_n)}{n-q} \right) \rightsquigarrow \mathcal{N}(0, 1).$$

Proof.

$$\frac{\mathbf{y}^{*\top} \mathbf{A}_n \mathbf{y}^*}{\mathbf{y}^{*\top} \mathbf{y}^*} - \frac{\text{tr}(\mathbf{A}_n)}{n-q} = \frac{(\phi^{1/2} \mathbf{y})^\top \left(\tilde{\mathbf{U}}_a \mathbf{A}_n \tilde{\mathbf{U}}_a^\top - \frac{\text{tr}(\mathbf{A}_n)}{n-q} \tilde{\mathbf{U}}_a \tilde{\mathbf{U}}_a^\top \right) (\phi^{1/2} \mathbf{y})}{\phi \mathbf{y}^{*\top} \mathbf{y}^*}.$$

$$\begin{aligned} & (\phi^{1/2} \mathbf{y})^\top \left(\tilde{\mathbf{U}}_a \mathbf{A}_n \tilde{\mathbf{U}}_a^\top - \frac{\text{tr}(\mathbf{A}_n)}{n-q} \tilde{\mathbf{U}}_a \tilde{\mathbf{U}}_a^\top \right) (\phi^{1/2} \mathbf{y}) \\ &= (\phi^{1/2} \varepsilon)^\top \left(\tilde{\mathbf{U}}_a \mathbf{A}_n \tilde{\mathbf{U}}_a^\top - \frac{\text{tr}(\mathbf{A}_n)}{n-q} \tilde{\mathbf{U}}_a \tilde{\mathbf{U}}_a^\top \right) (\phi^{1/2} \varepsilon) + \\ & \quad 2\phi^{1/2} (\phi^{1/2} \varepsilon)^\top \left(\tilde{\mathbf{U}}_a \mathbf{A}_n \tilde{\mathbf{U}}_a^\top - \frac{\text{tr}(\mathbf{A}_n)}{n-q} \tilde{\mathbf{U}}_a \tilde{\mathbf{U}}_a^\top \right) \mathbf{X}_b \boldsymbol{\beta}_b + \\ & \quad \phi \boldsymbol{\beta}_b^\top \mathbf{X}_b^\top \left(\tilde{\mathbf{U}}_a \mathbf{A}_n \tilde{\mathbf{U}}_a^\top - \frac{\text{tr}(\mathbf{A}_n)}{n-q} \tilde{\mathbf{U}}_a \tilde{\mathbf{U}}_a^\top \right) \mathbf{X}_b \boldsymbol{\beta}_b \end{aligned}$$

From (Jiang, 1996, Theorem 5.1),

□

As in Vershynin (2018), sub-gaussian norm of a sub-gaussian random variable is defined as

$$\|X\|_{\psi_2} = \inf\{t > 0 : \mathbb{E} \exp(X^2/t^2) \leq 2\}.$$

A random vector $Z \in \mathbb{R}^p$ is called sub-gaussian if $z^\top Z$ are sub-gaussian random variables for all $z \in \mathbb{R}^p$. The sub-gaussian norm of Z is defined as

$$\|Z\|_{\psi_2} = \sup_{z \in S^{p-1}} \|z^\top Z\|_{\psi_2},$$

where S^{p-1} is the unit sphere in \mathbb{R}^p .

Suppose $\mathbf{X}_b = \mathbf{Z}_b \Gamma + \mathbf{1}_n \mu_b^\top$, where the rows of \mathbf{Z}_b are iid sub-gaussian random vectors with identity covariance matrix.

The following lemma is a simple extension of Theorem 4.6.1 of Vershynin (2018).

Lemma 2. *Let \mathbf{Z} be an $N \times n$ random matrix whose columns Z_i are independent sub-gaussian random vectors with $\mathbb{E}(Z_i) = 0$, $\text{Var}(Z_i) = \mathbf{I}_n$. Suppose $K := \max_i \|Z_i\|_{\psi_2}$ is uniformly bounded. Write $Z_i = (z_{i1}, \dots, z_{iN})^\top$. Assume that $\mathbb{E}(z_{i\ell}^4) = 3 + \Delta < \infty$ and for any intergers $\ell_v \geq 0$ with $\sum_{v=1}^s \ell_v \leq 4$,*

$$\mathbb{E}(Z_{ij_1}^{\ell_1} Z_{ij_2}^{\ell_2} \cdots Z_{ij_s}^{\ell_s}) = \mathbb{E}(Z_{ij_1}^{\ell_1}) \mathbb{E}(Z_{ij_2}^{\ell_2}) \cdots \mathbb{E}(Z_{ij_s}^{\ell_s})$$

Let \mathbf{W} be a nonrandom $N \times N$ symmetric matrix. Then

$$\|\mathbf{Z}^\top \mathbf{W} \mathbf{Z} - \text{tr}(\mathbf{W}) \mathbf{I}_n\| = O_P(\sqrt{n} \|\mathbf{W}\|_F + n \|\mathbf{W}\|).$$

TO BE DONE:

$$\|U \mathbf{Z}^\top \mathbf{W} \mathbf{Z} U^\top - \text{tr}(\mathbf{W}) \mathbf{I}_n\|$$

Next we verify the following. Let $\Sigma = \Gamma^\top \Gamma$.

Proof. Let

$$\mathbf{B} = \mathbf{X}_b^* \mathbf{X}_b^{*\top} = \tilde{\mathbf{U}}_a^\top \mathbf{X}_b \mathbf{X}_b^\top \tilde{\mathbf{U}}_a = \tilde{\mathbf{U}}_a^\top \mathbf{Z}_b \Gamma \Gamma^\top \mathbf{Z}_b^\top \tilde{\mathbf{U}}_a$$

Note that Lemma 2 implies that

$$\|\mathbf{B} - \text{tr}(\Sigma) \mathbf{I}_{n-q}\| = O_P(\sqrt{n} \|\Sigma\|_F + n \|\Sigma\|).$$

That is, uniformly for $i = 1, \dots, n - q$,

$$\frac{\lambda_i(\mathbf{B})}{\text{tr}(\Sigma)} = 1 + O_P\left(\frac{\sqrt{n} \|\Sigma\|_F}{\text{tr}(\Sigma)} + \frac{n \|\Sigma\|}{\text{tr}(\Sigma)}\right)$$

Define

$$\delta_i = \frac{\lambda_i(\mathbf{B})}{\text{tr}(\Sigma)} - 1$$

$$\eta = \frac{\sqrt{n} \|\Sigma\|_F}{\text{tr}(\Sigma)} + \frac{n \|\Sigma\|}{\text{tr}(\Sigma)}$$

We assume $\eta \rightarrow 0$.

Thus, we need to verify

$$\tilde{\mathbf{U}}_a \mathbf{B}^{-1} \tilde{\mathbf{U}}_a^\top - \frac{\text{tr}(\mathbf{B}^{-1})}{n - q} \tilde{\mathbf{U}}_a \tilde{\mathbf{U}}_a^\top$$

satisfies that

$$\left\| \mathbf{B}^{-1} - \frac{\text{tr}(\mathbf{B}^{-1})}{n - q} \mathbf{I}_{n-q} \right\|^2 \bigg/ \text{tr} \left(\mathbf{B}^{-1} - \frac{\text{tr}(\mathbf{B}^{-1})}{n - q} \mathbf{I}_{n-q} \right)^2 \rightarrow 0. \quad (1)$$

Note that

$$\text{tr} \left(\mathbf{B}^{-1} - \frac{\text{tr}(\mathbf{B}^{-1})}{n-q} \mathbf{I}_{n-q} \right)^2 = \sum_{i=1}^{n-q} \frac{1}{\lambda_i^2(\mathbf{B})} - \frac{1}{n-q} \left(\sum_{i=1}^{n-q} \frac{1}{\lambda_i(\mathbf{B})} \right)^2.$$

By Taylor's theorem, uniformly for $i = 1, \dots, n$, we have

$$\begin{aligned} \frac{1}{\lambda_i(\mathbf{B})} &= \frac{1}{\text{tr}(\mathbf{\Sigma})} \frac{1}{1 + \delta_i} = \frac{1}{\text{tr}(\mathbf{\Sigma})} (1 - \delta_i + \delta_i^2 + O_P(\eta^3)), \\ \frac{1}{\lambda_i^2(\mathbf{B})} &= \frac{1}{\text{tr}^2(\mathbf{\Sigma})} \frac{1}{(1 + \delta_i)^2} = \frac{1}{\text{tr}^2(\mathbf{\Sigma})} (1 - 2\delta_i + 3\delta_i^2 + O_P(\eta^3)). \end{aligned}$$

Thus,

$$\begin{aligned} &\text{tr} \left(\mathbf{B}^{-1} - \frac{\text{tr}(\mathbf{B}^{-1})}{n-q} \mathbf{I}_{n-q} \right)^2 \\ &= \sum_{i=1}^{n-q} \frac{1}{\text{tr}^2(\mathbf{\Sigma})} (1 - 2\delta_i + 3\delta_i^2 + O_P(\eta^3)) - \frac{1}{n-q} \left(\sum_{i=1}^{n-q} \frac{1}{\text{tr}(\mathbf{\Sigma})} (1 - \delta_i + \delta_i^2 + O_P(\eta^3)) \right)^2 \\ &= \frac{1}{\text{tr}^2(\mathbf{\Sigma})} \left(n - q - 2 \sum_{i=1}^{n-q} \delta_i + 3 \sum_{i=1}^{n-q} \delta_i^2 + O_P(n\eta^3) - \frac{1}{n-q} \left(n - q - \sum_{i=1}^{n-q} \delta_i + \sum_{i=1}^{n-q} \delta_i^2 + O_P(n\eta^3) \right)^2 \right) \\ &= \frac{1}{\text{tr}^2(\mathbf{\Sigma})} \left(n - q - 2 \sum_{i=1}^{n-q} \delta_i + 3 \sum_{i=1}^{n-q} \delta_i^2 + O_P(n\eta^3) - (n-q) \left(1 - \frac{1}{n-q} \sum_{i=1}^{n-q} \delta_i + \frac{1}{n-q} \sum_{i=1}^{n-q} \delta_i^2 + O_P(\eta^3) \right)^2 \right) \\ &= \frac{1}{\text{tr}^2(\mathbf{\Sigma})} \left(n - q - 2 \sum_{i=1}^{n-q} \delta_i + 3 \sum_{i=1}^{n-q} \delta_i^2 + O_P(n\eta^3) \right. \\ &\quad \left. - (n-q) \left(1 + \left(\frac{1}{n-q} \sum_{i=1}^{n-q} \delta_i \right)^2 - \frac{2}{n-q} \sum_{i=1}^{n-q} \delta_i + \frac{2}{n-q} \sum_{i=1}^{n-q} \delta_i^2 + O_P(\eta^3) \right) \right) \\ &= \frac{1}{\text{tr}^2(\mathbf{\Sigma})} \left(\sum_{i=1}^{n-q} \delta_i^2 - \frac{1}{n-q} \left(\sum_{i=1}^{n-q} \delta_i \right)^2 + O_P(n\eta^3) \right) \\ &= \frac{1}{\text{tr}^2(\mathbf{\Sigma})} \left(\text{tr} \left(\frac{1}{\text{tr}(\mathbf{\Sigma})} \mathbf{B} - \mathbf{I}_{n-q} \right)^2 - \frac{1}{n-q} \left(\frac{\text{tr}(\mathbf{B})}{\text{tr}(\mathbf{\Sigma})} - (n-q) \right)^2 + O_P(n\eta^3) \right) \\ &= \frac{1}{\text{tr}^2(\mathbf{\Sigma})} \left(\frac{1}{\text{tr}^2(\mathbf{\Sigma})} \left(\text{tr}(\mathbf{B}^2) - \frac{1}{n-q} \text{tr}^2(\mathbf{B}) \right) + O_P(n\eta^3) \right). \end{aligned}$$

□

Lemma 3. Let \mathbf{Z} be an $n \times m$ random matrix whose rows Z_i are independent random vectors with $\mathbb{E}(Z_i) = 0$, $\text{Var}(Z_i) = \mathbf{I}_m$. Write $Z_i = (z_{i1}, \dots, z_{im})^\top$. Assume that $\mathbb{E}(z_{i\ell}^4) = 3 + \Delta < \infty$ and for any intergers $\ell_v \geq 0$ with $\sum_{v=1}^s \ell_v \leq 8$. Let \mathbf{B} be a $m \times m$ symmetric matrix. Let \mathbf{P} be a $n \times n$ symmetric maitrx. Then

$$\text{tr}(\mathbf{PZBZ}^\top \mathbf{P})^2 - \frac{1}{\text{tr}(\mathbf{P}^2)} \left(\text{tr}(\mathbf{PZBZ}^\top \mathbf{P}) \right)^2$$

Proof. Let $\mathbf{Q} = \mathbf{P}^2$,

$$\begin{aligned}\text{tr}(\mathbf{P}\mathbf{Z}\mathbf{B}\mathbf{Z}^\top\mathbf{P}) &= \text{tr}(\mathbf{Z}\mathbf{B}\mathbf{Z}^\top\mathbf{Q}) \\ &= \sum_{i=1}^n \sum_{j=1}^n Z_i^\top \mathbf{B} Z_j q_{j,i} \\ &= \sum_{i=1}^n q_{i,i} Z_i^\top \mathbf{B} Z_i + \sum_{i=1}^n \sum_{j \neq i}^n q_{i,j} Z_i^\top \mathbf{B} Z_j\end{aligned}$$

Then

$$\left(\text{tr}(\mathbf{P}\mathbf{Z}\mathbf{B}\mathbf{Z}^\top\mathbf{P})\right)^2 = \sum_{i=1}^n \sum_{i=1}^n q_{i,i} q_{j,j} Z_i^\top \mathbf{B} Z_i Z_j^\top \mathbf{B} Z_j + \sum_{i \neq j}^n \sum_{k \neq l}^n q_{i,j} q_{k,l} Z_i^\top \mathbf{B} Z_j Z_k^\top \mathbf{B} Z_l + \textcolor{red}{what?}$$

On the other hand,

$$\begin{aligned}\text{tr}(\mathbf{P}\mathbf{Z}\mathbf{B}\mathbf{Z}^\top\mathbf{P})^2 &= \text{tr}(\mathbf{Z}\mathbf{B}\mathbf{Z}^\top\mathbf{Q})^2 \\ &= \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n q_{j,k} q_{l,i} Z_i^\top \mathbf{B} Z_j Z_k^\top \mathbf{B} Z_l \\ &= \sum_{i \neq j}^n \sum_{k \neq l}^n q_{j,k} q_{l,i} Z_i^\top \mathbf{B} Z_j Z_k^\top \mathbf{B} Z_l + 2 \sum_{i=1}^n \sum_{j \neq k}^n q_{i,j} q_{i,k} Z_i^\top \mathbf{B} Z_j Z_j^\top \mathbf{B} Z_k + \sum_{i=1}^n \sum_{j=1}^n q_{i,j}^2 Z_i^\top \mathbf{B} Z_i Z_j^\top \mathbf{B} Z_j.\end{aligned}$$

Thus,

$$\begin{aligned}\text{tr}(\mathbf{P}\mathbf{Z}\mathbf{B}\mathbf{Z}^\top\mathbf{P})^2 - \frac{1}{\text{tr}(\mathbf{P}^2)} \left(\text{tr}(\mathbf{P}\mathbf{Z}\mathbf{B}\mathbf{Z}^\top\mathbf{P})\right)^2 &= \sum_{i \neq j}^n \sum_{k \neq l}^n (q_{i,l} q_{j,k} - C q_{i,j} q_{k,l}) Z_i^\top \mathbf{B} Z_j Z_k^\top \mathbf{B} Z_l + 2 \sum_{i=1}^n \sum_{j \neq k}^n q_{i,j} q_{i,k} Z_i^\top \mathbf{B} Z_i Z_j^\top \mathbf{B} Z_k \\ &\quad + \sum_{i=1}^n \sum_{j=1}^n (q_{i,j}^2 - C q_{i,i} q_{j,j}) Z_i^\top \mathbf{B} Z_i Z_j^\top \mathbf{B} Z_j \\ &=: A_1 + A_2 + A_3\end{aligned}$$

en

$$\begin{aligned}A_1 &= \sum_{i \neq j}^n \sum_{k=i, l=j}^n + \sum_{i \neq j}^n \sum_{k=j, l=i}^n + \sum_{i \neq j}^n \sum_{k=i, l \notin \{i,j\}}^n + \sum_{i \neq j}^n \sum_{k=j, l \notin \{i,j\}}^n + \sum_{i \neq j}^n \sum_{l=i, k \notin \{i,j\}}^n + \sum_{i \neq j}^n \sum_{l=j, k \notin \{i,j\}}^n + \sum_{i \neq j}^n \sum_{k \notin \{i,j\}, l \notin \{i,j\}, k \neq l}^n \\ &=: A_{1,1} + \dots + A_{1,7}\end{aligned}$$

Note that

$$A_1^{(1)} := A_{1,1} + A_{1,2} = \sum_{i \neq j} (q_{i,i} q_{j,j} + (1 - 2C) q_{i,j}^2) (Z_i^\top \mathbf{B} Z_j)^2$$

and

$$A_1^{(2)} := A_{1,3} + A_{1,4} + A_{1,5} + A_{1,6} = \sum_{i,j,k \text{ distinct}} (2q_{i,i}q_{j,k} - (2 - 4C)q_{i,j}q_{i,k}) Z_i^\top \mathbf{B} Z_j Z_i^\top \mathbf{B} Z_k$$

and

$$A_1^{(3)} := A_{1,7} = \sum_{i,j,k,l \text{ distinct}} (q_{i,l}q_{j,k} - Cq_{i,j}q_{k,l}) Z_i^\top \mathbf{B} Z_j Z_k^\top \mathbf{B} Z_l$$

For $A_1^{(3)}$, we have,

$$\mathbb{E} A_1^{(3)} = 0$$

and

$$\begin{aligned} & \mathbb{E}(A_1^{(3)})^2 \\ &= \mathbb{E} \sum_{i,j,k,l \text{ distinct } \sigma: \text{permutation of } i,j,k,l} \sum_{\sigma}^{24} \\ & \quad (q_{i,l}q_{j,k} - Cq_{i,j}q_{k,l}) Z_i^\top \mathbf{B} Z_j Z_k^\top \mathbf{B} Z_l (q_{\sigma(i),\sigma(l)}q_{\sigma(j),\sigma(k)} - Cq_{\sigma(i),\sigma(j)}q_{\sigma(k),\sigma(l)}) Z_{\sigma(i)}^\top \mathbf{B} Z_{\sigma(j)} Z_{\sigma(k)}^\top \mathbf{B} Z_{\sigma(l)} \\ &\leq \sum_{i,j,k,l \text{ distinct } \sigma: \text{permutation of } i,j,k,l} \sum_{\sigma}^{24} \\ & \quad \left| (q_{i,l}q_{j,k} - Cq_{i,j}q_{k,l}) (q_{\sigma(i),\sigma(l)}q_{\sigma(j),\sigma(k)} - Cq_{\sigma(i),\sigma(j)}q_{\sigma(k),\sigma(l)}) \right| \\ & \quad \mathbb{E} \left| Z_i^\top \mathbf{B} Z_j Z_k^\top \mathbf{B} Z_l Z_{\sigma(i)}^\top \mathbf{B} Z_{\sigma(j)} Z_{\sigma(k)}^\top \mathbf{B} Z_{\sigma(l)} \right| \\ &\leq \sum_{i,j,k,l \text{ distinct } \sigma: \text{permutation of } i,j,k,l} \sum_{\sigma}^{24} \\ & \quad (q_{i,l}q_{j,k} - Cq_{i,j}q_{k,l})^2 + (q_{\sigma(i),\sigma(l)}q_{\sigma(j),\sigma(k)} - Cq_{\sigma(i),\sigma(j)}q_{\sigma(k),\sigma(l)})^2 \\ & \quad \sqrt{\mathbb{E} (Z_i^\top \mathbf{B} Z_j Z_k^\top \mathbf{B} Z_l)^2 \mathbb{E} (Z_{\sigma(i)}^\top \mathbf{B} Z_{\sigma(j)} Z_{\sigma(k)}^\top \mathbf{B} Z_{\sigma(l)})^2} \\ &= 48 \sum_{i,j,k,l \text{ distinct}} (q_{i,l}q_{j,k} - Cq_{i,j}q_{k,l})^2 \text{tr}^2(\mathbf{B}^2) \\ &\leq 48 \sum_{i,j,k,l \text{ distinct}} (q_{i,l}q_{j,k} - Cq_{i,j}q_{k,l})^2 \text{tr}^2(\mathbf{B}^2) \\ &\leq 96(1 + C^2) \sum_{i,j,k,l \text{ distinct}} (q_{i,l}q_{j,k})^2 \text{tr}^2(\mathbf{B}^2) \\ &\leq 96(1 + C^2) \sum_{i \neq j} \sum_{k \neq l} (q_{i,j}q_{k,l})^2 \text{tr}^2(\mathbf{B}^2) \\ &\leq 96(1 + C^2) \left(\sum_{i \neq j} (q_{i,j})^2 \right)^2 \text{tr}^2(\mathbf{B}^2) \end{aligned}$$

For $A_1^{(2)}$, we have $E A_1^{(2)} = 0$, and

$$\begin{aligned}
& E \left(A_1^{(2)} \right)^2 \\
&= \sum_{i,j,k \text{ distinct}} \sum_{i',j',k' \text{ distinct}} \left(2q_{i,i}q_{j,k} - (2-4C)q_{i,j}q_{i,k} \right) Z_i^\top \mathbf{B} Z_j Z_i^\top \mathbf{B} Z_k \left(2q_{i',i'}q_{j',k'} - (2-4C)q_{i',j'}q_{i',k'} \right) Z_{i'}^\top \mathbf{B} Z_{j'} Z_{i'}^\top \mathbf{B} Z_{k'} \\
&= \sum_{i,j,k \text{ distinct}} \left(\sum_{\{i',j',k'\}=\{i,j,k\}} + \sum_{\{j',k'\}=\{j,k\}, i' \notin \{j,k\}} \right)
\end{aligned}$$

Note that

$$\begin{aligned}
& \sum_{i,j,k \text{ distinct}} \sum_{\{i',j',k'\}=\{i,j,k\}} \\
&\leq \sum_{i,j,k \text{ distinct}} \sum_{\{i',j',k'\}=\{i,j,k\}} \left((2q_{i,i}q_{j,k} - (2-4C)q_{i,j}q_{i,k})^2 + (2q_{i',i'}q_{j',k'} - (2-4C)q_{i',j'}q_{i',k'})^2 \right) E(Z_1^\top \mathbf{B} Z_2)^4 \\
&= 12 \sum_{i,j,k \text{ distinct}} (2q_{i,i}q_{j,k} - (2-4C)q_{i,j}q_{i,k})^2 E(Z_1^\top \mathbf{B} Z_2)^4 \\
&\leq 96 \sum_{i,j,k \text{ distinct}} \left((q_{i,i}q_{j,k})^2 + (1-2C)^2 (q_{i,j}q_{i,k})^2 \right) E(Z_1^\top \mathbf{B} Z_2)^4 \\
&= O(1) \left(\left(\sum_{i=1}^n q_{i,i}^2 \right) \left(\sum_{i \neq j}^n q_{i,j}^2 \right) + (1-2C)^2 \sum_{i=1}^n \left(\sum_{j:j \neq i} q_{i,j}^2 \right)^2 \right) \text{tr}^2(\mathbf{B}^2)
\end{aligned}$$

Now we deal with $A_1^{(1)}$.

$$A_1^{(1)} = \sum_{i \neq j} (q_{i,i}q_{j,j} + (1-2C)q_{i,j}^2) (Z_i^\top \mathbf{B} Z_j)^2$$

$$\begin{aligned}
E A_1^{(1)} &= \sum_{i \neq j} q_{i,i}q_{j,j} \text{tr}(\mathbf{B}^2) + \sum_{i \neq j} (1-2C)q_{i,j}^2 \text{tr}(\mathbf{B}^2) \\
&= \left(\left(\sum_{i=1}^n q_{i,i} \right)^2 - \sum_{i=1}^n q_{i,i}^2 \right) \text{tr}(\mathbf{B}^2) + \sum_{i \neq j} (1-2C)q_{i,j}^2 \text{tr}(\mathbf{B}^2)
\end{aligned}$$

And

$$\begin{aligned}
& \text{Var} \left(A_1^{(1)} \right) \\
&= \mathbb{E} \left(\sum_{i \neq j} (q_{i,i} q_{j,j} + (1 - 2C) q_{i,j}^2) \left((Z_i^\top \mathbf{B} Z_j)^2 - \text{tr}(\mathbf{B}^2) \right) \right)^2 \\
&= 2 \sum_{i \neq j} (q_{i,i} q_{j,j} + (1 - 2C) q_{i,j}^2)^2 \mathbb{E} \left((Z_i^\top \mathbf{B} Z_j)^2 - \text{tr}(\mathbf{B}^2) \right)^2 \\
&\quad + 4 \sum_{i,j,k}^* (q_{i,i} q_{j,j} + (1 - 2C) q_{i,j}^2) (q_{i,i} q_{k,k} + (1 - 2C) q_{i,k}^2) \mathbb{E} \left((Z_i^\top \mathbf{B} Z_j)^2 - \text{tr}(\mathbf{B}^2) \right) \left((Z_i^\top \mathbf{B} Z_k)^2 - \text{tr}(\mathbf{B}^2) \right) \\
&= 2 \sum_{i \neq j} (q_{i,i} q_{j,j} + (1 - 2C) q_{i,j}^2)^2 \mathbb{E} \left((Z_i^\top \mathbf{B} Z_j)^2 - \text{tr}(\mathbf{B}^2) \right)^2 \\
&\quad + 4 \sum_{i,j,k}^* (q_{i,i} q_{j,j} + (1 - 2C) q_{i,j}^2) (q_{i,i} q_{k,k} + (1 - 2C) q_{i,k}^2) \mathbb{E} \left(Z_i^\top \mathbf{B}^2 Z_i - \text{tr}(\mathbf{B}^2) \right)^2
\end{aligned}$$

But

$$\mathbb{E} \left((Z_i^\top \mathbf{B} Z_j)^2 - \text{tr}(\mathbf{B}^2) \right)^2 = \mathbb{E} (Z_i^\top \mathbf{B} Z_j)^4 - \text{tr}^2(\mathbf{B}^2) = O(\text{tr}^2(\mathbf{B}^2))$$

and

$$\mathbb{E} \left(Z_i^\top \mathbf{B}^2 Z_i - \text{tr}(\mathbf{B}^2) \right)^2 \leq \sqrt{\mathbb{E} (Z_i^\top \mathbf{B}^2 Z_i - \text{tr}(\mathbf{B}^2))^4} = O(\text{tr}(\mathbf{B}^4)).$$

Thus,

$$\begin{aligned}
& \text{Var} \left(A_1^{(1)} \right) \\
&= O(\text{tr}^2(\mathbf{B}^2)) \left(\left(\sum_{i=1}^n q_{i,i}^2 \right)^2 + (1 - 2C)^2 \sum_{i \neq j} q_{i,j}^4 \right) \\
&\quad + O(\text{tr}(\mathbf{B}^4)) \left(\left(\sum_{i=1}^n q_{i,i}^2 \right) \left(\sum_{i=1}^n q_{i,i} \right)^2 + \left(\sum_{i=1}^n q_{i,i} \right) \left(\sum_{i=1}^n \sum_{j=1}^n q_{i,i} q_{i,j}^2 \right) + \sum_{i=1}^n \left(\sum_{j=1}^n q_{i,j}^2 \right)^2 \right)
\end{aligned}$$

Now we deal with A_2 .

$$\begin{aligned}
A_2 &= \sum_{i=1}^n \sum_{j \neq k}^n q_{i,j} q_{i,k} Z_i^\top \mathbf{B} Z_i Z_j^\top \mathbf{B} Z_k \\
&= 2 \sum_{i,j}^* q_{i,j} q_{i,i} Z_i^\top \mathbf{B} Z_i Z_i^\top \mathbf{B} Z_j + \sum_{i,j,k}^* q_{i,j} q_{i,k} Z_i^\top \mathbf{B} Z_i Z_j^\top \mathbf{B} Z_k
\end{aligned}$$

□

Appendices

Appendix A haha1

Proof of Proposition 1. We assume $0 < \alpha < 1$ since the case $\alpha = 0$ or 1 is trivial. Note that the condition implies $\int [\varphi(\mathbf{y}) - \alpha] \mathcal{N}_n(0, \phi^{-1} \mathbf{I}_n)(d\mathbf{y}) = 0$. Hence it suffices to prove $\varphi(\mathbf{y}) \geq \alpha$, a.s. We prove this by contradiction. Suppose $\lambda(\{\mathbf{y} : \varphi(\mathbf{y}) < \alpha\}) > 0$. Then there exists a $\eta > 0$, such that $\lambda(\{\mathbf{y} : \varphi(\mathbf{y}) < \alpha - \eta\}) > 0$. We denote $E = \{\mathbf{y} : \varphi(\mathbf{y}) < \alpha - \eta\}$. From Lebesgue density theorem (Cohn, 2013, Corollary 6.2.6), there exists a point $z \in E$, such that, for each $\epsilon > 0$ there is a $\delta_\epsilon > 0$ such that

$$\left| \frac{\lambda(E^\complement \cap C_\epsilon)}{\lambda(C_\epsilon)} \right| < \epsilon,$$

where $C_\epsilon = \prod_{i=1}^n [z_i - \delta_\epsilon, z_i + \delta_\epsilon]$. We put

$$\epsilon = \left(\frac{\sqrt{\pi}}{\sqrt{2}\Phi^{-1}\left(1 - \frac{\eta}{6n}\right)} \right)^n \frac{\eta}{3}.$$

Then for any $\phi > 0$,

$$\begin{aligned} \alpha &\leq \int_{\mathbb{R}^n} \varphi(\mathbf{y}) \mathcal{N}_n(z, \phi^{-1} \mathbf{I}_n)(d\mathbf{y}) \\ &= \int_{E \cap C_\epsilon} \varphi(\mathbf{y}) \mathcal{N}_n(z, \phi^{-1} \mathbf{I}_n)(d\mathbf{y}) + \int_{E^\complement \cap C_\epsilon} \varphi(\mathbf{y}) \mathcal{N}_n(z, \phi^{-1} \mathbf{I}_n)(d\mathbf{y}) + \int_{C_\epsilon^\complement} \varphi(\mathbf{y}) \mathcal{N}_n(z, \phi^{-1} \mathbf{I}_n)(d\mathbf{y}) \\ &\leq \alpha - \eta + \int_{E^\complement \cap C_\epsilon} \mathcal{N}_n(z, \phi^{-1} \mathbf{I}_n)(d\mathbf{y}) + \int_{C_\epsilon^\complement} \mathcal{N}_n(z, \phi^{-1} \mathbf{I}_n)(d\mathbf{y}) \\ &\leq \alpha - \eta + \left(\frac{\phi}{2\pi} \right)^{n/2} \lambda(E^\complement \cap C_\epsilon) + 2n \left(1 - \Phi(\sqrt{\phi}\delta_\epsilon) \right) \\ &\leq \alpha - \eta + \left(\frac{\phi}{2\pi} \right)^{n/2} \epsilon (2\delta_\epsilon)^n + 2n \left(1 - \Phi(\sqrt{\phi}\delta_\epsilon) \right) \\ &= \alpha - \eta + \left(\frac{\sqrt{\phi}\delta_\epsilon}{\Phi^{-1}\left(1 - \frac{\eta}{6n}\right)} \right)^n \frac{\eta}{3} + 2n \left(1 - \Phi(\sqrt{\phi}\delta_\epsilon) \right). \end{aligned}$$

Putting

$$\phi = \left(\frac{\Phi^{-1}\left(1 - \frac{\eta}{6n}\right)}{\delta_\epsilon} \right)^2$$

yields the contradiction $\alpha \leq \alpha - (2/3)\eta$. This completes the proof. \square

Proof of Proposition 2. For positive integer m , define $[m] = \{1, \dots, m\}$. For a set A , denote

by $|A|$ its cardinality. We have

$$\begin{aligned} k_\kappa &= \left| \left\{ i \in [n-q] : \frac{\gamma_i^2}{\gamma_i + \kappa} - \frac{1}{n-q} \sum_{j=1}^{n-q} \frac{\gamma_j \gamma_i}{\gamma_j + \kappa} > 0 \right\} \right| \\ &= \left| \left\{ i \in [n-q] : \frac{\gamma_i}{\gamma_i + \kappa} > \frac{1}{n-q} \sum_{j=1}^{n-q} \frac{\gamma_j}{\gamma_j + \kappa} \right\} \right|. \end{aligned}$$

Let X be a random variable uniformly distributed on $\{\gamma_1, \dots, \gamma_{n-q}\}$. That is, $\Pr(X = \gamma_i) = 1/(n-q)$, $i = 1, \dots, n-q$. Then it can be seen that

$$k_\kappa = (n-q) \Pr \left(\frac{X}{X + \kappa} > \mathbb{E} \left[\frac{X}{X + \kappa} \right] \right).$$

Hence we only need to verify

$$\Pr \left(\frac{X}{X + \kappa_1} > \mathbb{E} \left[\frac{X}{X + \kappa_1} \right] \right) \geq \Pr \left(\frac{X}{X + \kappa_2} > \mathbb{E} \left[\frac{X}{X + \kappa_2} \right] \right). \quad (2)$$

Let $Y = X/(X + \kappa_2)$. Then

$$\frac{X}{(X + \kappa_1)} = \frac{\kappa_2 Y}{\kappa_1 + (\kappa_2 - \kappa_1)Y} := f(Y).$$

Note that $f(Y)$ is increasing for $Y \geq 0$. Then the inequality (2) is equivalent to

$$\Pr(Y > f^{-1}(\mathbb{E} f(Y))) \geq \Pr(Y > \mathbb{E} Y).$$

Hence we only need to verify $f^{-1}(\mathbb{E} f(Y)) \leq \mathbb{E} Y$, or equivalently, $\mathbb{E} f(Y) \leq f(\mathbb{E} Y)$. But the last inequality is a direct consequence of the concavity of $f(Y)$. This completes the proof. \square

Lemma 4. *Let \mathbf{W} be an $N \times N$ positive semi-definite matrix. Let Z be an N dimensional sub-gaussian random vector with $\mathbb{E} Z = 0$, $\text{Var}(Z) = \mathbf{I}_N$ and $\|Z\|_{\psi_2} \leq K$. For all $t > 0$,*

$$\Pr \left\{ Z^\top \mathbf{W} Z > \right\} \leq e^{-t}.$$

Remark 1. This lemma is adapted from Hsu et al. (2012) Theorem 2.1 with minor modifications. Indeed, their result did not track the variance of Z .

Proof. Let $\mathbf{A} = \sqrt{\mathbf{W}}$. Let Z^* be a vector of N independent standard Gaussian random variables which are independent of Z .

$$\mathbb{E}[\exp(\lambda Z^{*\top} \mathbf{A} Z)] = \mathbb{E} \mathbb{E}[\exp(\lambda Z^{*\top} \mathbf{A} Z) | Z] = \mathbb{E}[\exp(\frac{\lambda^2}{2} Z^\top \mathbf{W} Z)]$$

\square

Lemma 5.

Proof.

□

Proof of Lemma 2.

$$\|\mathbf{Z}^\top \mathbf{W} \mathbf{Z} - \text{tr}(\mathbf{W}) \mathbf{I}_n\| \leq \|\mathbf{Z}^\top \mathbf{W} \mathbf{Z} - \text{diag}(\mathbf{Z}^\top \mathbf{W} \mathbf{Z})\| + \|\text{diag}(\mathbf{Z}^\top \mathbf{W} \mathbf{Z}) - \text{tr}(\mathbf{W}) \mathbf{I}_n\|$$

We have

$$\|\text{diag}(\mathbf{Z}^\top \mathbf{W} \mathbf{Z}) - \text{tr}(\mathbf{W}) \mathbf{I}_n\| = \max_{1 \leq i \leq n} |Z_i^\top \mathbf{W} Z_i - \text{tr}(\mathbf{W})| \leq \sqrt{\sum_{i=1}^n (Z_i^\top \mathbf{W} Z_i - \text{tr}(\mathbf{W}))^2}$$

From (Chen et al., 2010, Proposition A.1),

$$\mathbb{E} \left[\sum_{i=1}^n \left(Z_i^\top \mathbf{W} Z_i - \text{tr}(\mathbf{W}) \right)^2 \right] = 2n \text{tr}(\mathbf{W}^2) + \Delta n \text{tr}(\mathbf{W} \circ \mathbf{W}) \leq (2 + \Delta) n \text{tr}(\mathbf{W}^2).$$

Hence

$$\|\text{diag}(\mathbf{Z}^\top \mathbf{W} \mathbf{Z}) - \text{tr}(\mathbf{W}) \mathbf{I}_n\| = O_P(\sqrt{n} \|\mathbf{W}\|_F).$$

Next we deal with

$$\|\mathbf{Z}^\top \mathbf{W} \mathbf{Z} - \text{diag}(\mathbf{Z}^\top \mathbf{W} \mathbf{Z})\|$$

From (Vershynin, 2018, Lemma 5.2), there is a $1/4$ -net \mathcal{C} of the unit sphere S^{n-1} such that $|\mathcal{C}| \leq 9^n$.

By (Vershynin, 2018, Exercise 4.4.3),

$$\|\mathbf{Z}^\top \mathbf{W} \mathbf{Z} - \text{diag}(\mathbf{Z}^\top \mathbf{W} \mathbf{Z})\| \leq 2 \sup_{x \in \mathcal{C}} \left| x^\top \left(\mathbf{Z}^\top \mathbf{W} \mathbf{Z} - \text{diag}(\mathbf{Z}^\top \mathbf{W} \mathbf{Z}) \right) x \right|.$$

Fix $x \in \mathcal{C}$. Then

$$\left| x^\top \left(\mathbf{Z}^\top \mathbf{W} \mathbf{Z} - \text{diag}(\mathbf{Z}^\top \mathbf{W} \mathbf{Z}) \right) x \right| = \left| \sum_{i=1}^n \sum_{j \neq i}^n x_i x_j Z_i^\top \mathbf{W} Z_j \right|$$

Now we bound the moment generating function of $\sum_{i=1}^n \sum_{j \neq i}^n x_i x_j Z_i^\top \mathbf{W} Z_j$. We apply the decoupling technique in Vershynin (2018), Section 6.1. Let $\delta_1, \dots, \delta_n$ be independent Bernoulli random variables with $\Pr\{\delta_i = 0\} = \Pr\{\delta_i = 1\} = 1/2$. For any $\lambda \in \mathbb{R}$,

$$\begin{aligned} \mathbb{E} \exp \left\{ \lambda \sum_{i=1}^n \sum_{j \neq i}^n x_i x_j Z_i^\top \mathbf{W} Z_j \right\} &= \mathbb{E} \exp \left\{ \mathbb{E} \left(4\lambda \sum_{i=1}^n \sum_{j=1}^n \delta_i (1 - \delta_j) x_i x_j Z_i^\top \mathbf{W} Z_j \middle| \mathbf{Z} \right) \right\} \\ &\leq \mathbb{E} \exp \left\{ 4\lambda \sum_{i=1}^n \sum_{j=1}^n \delta_i (1 - \delta_j) x_i x_j Z_i^\top \mathbf{W} Z_j \right\} \\ &= \mathbb{E} \exp \left\{ 4\lambda \left(\sum_{i: \delta_i=1} x_i Z_i \right)^\top \mathbf{W} \left(\sum_{j: \delta_j=0} x_j Z_j \right) \right\} \\ &\leq \max_{I \subset [n]} \mathbb{E} \exp \left\{ 4\lambda \left(\sum_{i \in I} x_i Z_i \right)^\top \mathbf{W} \left(\sum_{j \notin I} x_j Z_j \right) \right\}, \end{aligned}$$

where the first inequality follows from Jensen's inequality. Fix an $I \subset [n]$. From Vershynin (2018), Proposition 2.6.1, $\|\sum_{i \in I} x_i Z_i\|_{\psi_2} \leq C_1 K$, $\|\sum_{j \notin I} x_j Z_j\|_{\psi_2} \leq C_1 K$ for some absolute constant C_1 . Then Vershynin (2018), Lemma 6.2.2 and Lemma 6.2.3 imply that there exist absolute constants C_2, C_3 such that,

$$\mathbb{E} \exp \left\{ 4\lambda \left(\sum_{i \in I} x_i Z_i \right)^\top \mathbf{W} \left(\sum_{j \notin I} x_j Z_j \right) \right\} \leq \exp \{ C_2 K^4 \|\mathbf{W}\|_F^2 \lambda^2 \}$$

for all $|\lambda| \leq C_3/(K^2 \|\mathbf{W}\|)$. Note that this bound does not depend on $I \subset [n]$. It follows that

$$\mathbb{E} \exp \left\{ \lambda \sum_{i=1}^n \sum_{j \neq i}^n x_i x_j Z_i^\top \mathbf{W} Z_j \right\} \leq \exp \{ C_2 K^4 \|\mathbf{W}\|_F^2 \lambda^2 \},$$

for all $|\lambda| \leq C_3/(K^2 \|\mathbf{W}\|)$. Then applying Chernoff bound yields that, for any $t > 0$,

$$\begin{aligned} \Pr \left(\left| \sum_{i=1}^n \sum_{j \neq i}^n x_i x_j Z_i^\top \mathbf{W} Z_j \right| > t \right) &\leq \inf_{0 < \lambda \leq \frac{C_3}{K^2 \|\mathbf{W}\|}} 2 \exp \{ -\lambda t + C_2 K^4 \|\mathbf{W}\|_F^2 \lambda^2 \} \\ &\leq 2 \exp \left\{ -\min \left(\frac{t^2}{4C_2 K^4 \|\mathbf{W}\|_F^2}, \frac{C_3 t}{2K^2 \|\mathbf{W}\|} \right) \right\}. \end{aligned}$$

This inequality, combined with union bound, yields

$$\begin{aligned} \Pr \left(\|\mathbf{Z}^\top \mathbf{W} \mathbf{Z} - \text{diag}(\mathbf{Z}^\top \mathbf{W} \mathbf{Z})\| > t \right) &\leq \Pr \left(2 \sup_{x \in \mathcal{C}} \left| x^\top (\mathbf{Z}^\top \mathbf{W} \mathbf{Z} - \text{diag}(\mathbf{Z}^\top \mathbf{W} \mathbf{Z})) x \right| > t \right) \\ &\leq 2 \cdot 9^n \exp \left\{ -\min \left(\frac{t^2}{16C_2 K^4 \|\mathbf{W}\|_F^2}, \frac{C_3 t}{4K^2 \|\mathbf{W}\|} \right) \right\}. \end{aligned}$$

Thus, there exists a large $C > 0$ such that for every $t > 0$,

$$\Pr \left(\|\mathbf{Z}^\top \mathbf{W} \mathbf{Z} - \text{diag}(\mathbf{Z}^\top \mathbf{W} \mathbf{Z})\| > C(K^2(\sqrt{n} + t)\|\mathbf{W}\|_F + K^2(n + t^2)\|\mathbf{W}\|) \right) \leq 2 \exp\{-t^2\}.$$

Consequently, $\|\mathbf{Z}^\top \mathbf{W} \mathbf{Z} - \text{diag}(\mathbf{Z}^\top \mathbf{W} \mathbf{Z})\| = O_P(K^2(\sqrt{n}\|\mathbf{W}\|_F + n\|\mathbf{W}\|))$. This completes the proof. \square

Appendix B haha2

Theorem 2. Let ζ_1, \dots, ζ_d be iid random variables with mean 0 and variance 1, and assume $\mu_k := \mathbb{E}(\zeta_1^k)$ is finite for $k \leq 8$. Let $\boldsymbol{\zeta} = (\zeta_1, \dots, \zeta_d)^\top \in \mathbb{R}^d$. For $k = 1, \dots, K$, let $\mathbf{Q}_k = (q_{ij}^{(k)})$ be a $d \times d$ symmetric matrix and let $\check{\mathbf{Q}}_k = \text{diag}(q_{11}^{(k)}, \dots, q_{dd}^{(k)})$, $\hat{\mathbf{Q}}_k = \mathbf{I}_d - \check{\mathbf{Q}}_k$. Define $\hat{w}_k = \boldsymbol{\zeta}^\top \hat{\mathbf{Q}}_k \boldsymbol{\zeta}$, $\check{w}_k = \boldsymbol{\zeta}^\top \check{\mathbf{Q}}_k \boldsymbol{\zeta} - \text{tr}(\mathbf{Q}_k)$, and

$$W = \begin{pmatrix} \hat{w}_1 \\ \check{w}_1 \\ \vdots \\ \hat{w}_K \\ \check{w}_K \end{pmatrix} = \begin{pmatrix} \boldsymbol{\zeta}^\top \hat{\mathbf{Q}}_1 \boldsymbol{\zeta} \\ \boldsymbol{\zeta}^\top \check{\mathbf{Q}}_1 \boldsymbol{\zeta} - \text{tr}(\mathbf{Q}_1) \\ \vdots \\ \boldsymbol{\zeta}^\top \hat{\mathbf{Q}}_K \boldsymbol{\zeta} \\ \boldsymbol{\zeta}^\top \check{\mathbf{Q}}_K \boldsymbol{\zeta} - \text{tr}(\mathbf{Q}_K) \end{pmatrix} \in \mathbb{R}^{2K}.$$

Finally, let $Z \sim \mathcal{N}_{2K}(0, \mathbf{I}_{2K})$ and $\mathbf{V} = \text{Cov}(W)$. There is an absolute constant $0 < C < \infty$ such that

haha

Proof. Let $f : \mathbb{R}^{2K} \rightarrow \mathbb{R}$ be a four-times differentiable function. From xxx, there is a 4 - times differentiable function $g : \mathbb{R}^{2K} \rightarrow \mathbb{R}$ satisfying the Stein identity

$$\mathbb{E}[f(W)] - \mathbb{E}[f(\mathbf{V}^{1/2}W)] = \mathbb{E}[\nabla^\top \mathbf{V} \nabla g(W) - W^\top \nabla g(W)]$$

and

$$\left| \frac{\partial^k g(\mathbf{x})}{\prod_{j=1}^k \partial x_{i_j}} \right| \leq \frac{1}{k} \left| \frac{\partial^k f(\mathbf{x})}{\prod_{j=1}^k \partial x_{i_j}} \right| \quad \text{for all } \mathbf{x} = (x_1, \dots, x_{2K})^\top \in \mathbb{R}^{2K}, k = 1, 2, 3, \text{ and } i_j \in \{1, \dots, 2K\}.$$

To prove the theorem, we bound

$$S = \mathbb{E}[\nabla^\top \mathbf{V} \nabla g(W) - W^\top \nabla g(W)].$$

Next, we use exchangeability. Let $\zeta' = (\zeta'_1, \dots, \zeta'_d)^\top$ be an independent copy of ζ , and let $\underline{i} \in \{1, \dots, d\}$ be an independent and uniformly distributed random index. Define the vector $W' \in \mathbb{R}^{2K}$ exactly as we defined W , except that $\zeta_{\underline{i}}$ is replaced with $\zeta'_{\underline{i}}$ throughout. More precisely, let $e_i \in \mathbb{R}^d$ be the i th standard basis vector in \mathcal{R}^d and define

$$\begin{aligned} \hat{w}'_k &= (\zeta + (\zeta'_{\underline{i}} - \zeta_{\underline{i}})e_{\underline{i}})^\top \hat{\mathbf{Q}}_k (\zeta + (\zeta'_{\underline{i}} - \zeta_{\underline{i}})e_{\underline{i}}) \\ &= \hat{w}_k + 2(\zeta'_{\underline{i}} - \zeta_{\underline{i}})e_{\underline{i}}^\top \hat{\mathbf{Q}}_k \zeta, \end{aligned}$$

$$\begin{aligned} \check{w}'_k &= (\zeta + (\zeta'_{\underline{i}} - \zeta_{\underline{i}})e_{\underline{i}})^\top \check{\mathbf{Q}}_k (\zeta + (\zeta'_{\underline{i}} - \zeta_{\underline{i}})e_{\underline{i}}) - \text{tr}(\mathbf{Q}_k) \\ &= \check{w}_k + e_{\underline{i}}^\top \check{\mathbf{Q}}_k e_{\underline{i}} ((\zeta'_{\underline{i}})^2 - \zeta_{\underline{i}}^2), \end{aligned}$$

for $k = 1, \dots, K$. Then $W' = (\hat{w}'_1, \check{w}'_1, \dots, \hat{w}'_K, \check{w}'_K)^\top \in \mathbb{R}^{2K}$. Its straightforward to verify that

$$\mathbb{E}(\hat{w}'_k - \hat{w}_k | \zeta) = -\frac{2}{d} \hat{w}_k, \quad \mathbb{E}(\check{w}'_k - \check{w}_k | \zeta) = -\frac{1}{d} \check{w}_k.$$

Then

$$\mathbb{E}(W' - W | \zeta) = -\Lambda_K W,$$

where

$$\Lambda_1 = \begin{pmatrix} \frac{2}{d} & 0 \\ 0 & \frac{1}{d} \end{pmatrix}, \quad \Lambda_K = \begin{pmatrix} \Lambda_1 & 0 & \cdots & 0 \\ 0 & \Lambda_1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \Lambda_1 \end{pmatrix} \in \mathbb{R}^{2K \times 2K}.$$

By exchangeability, we have

$$\begin{aligned} 0 &= \frac{1}{2} \mathbb{E}[(W' - W)^\top \Lambda_K^{-\top} (\nabla g(W') + \nabla g(W))] \\ &= \mathbb{E}[(W' - W)^\top \Lambda_K^{-\top} \nabla g(W)] + \frac{1}{2} \mathbb{E}[(W' - W)^\top \Lambda_K^{-\top} (\nabla g(W') - \nabla g(W))] \\ &= -\mathbb{E}[W^\top \nabla g(W)] + \frac{1}{2} \mathbb{E}[(W' - W)^\top \Lambda_K^{-\top} (\nabla g(W') - \nabla g(W))]. \end{aligned}$$

That is,

$$\mathbb{E}[W^\top \nabla g(W)] = \frac{1}{2} \mathbb{E}[(W' - W)^\top \Lambda_K^{-\top} (\nabla g(W') - \nabla g(W))].$$

Apply Taylor's theorem,

$$\begin{aligned} & W^\top \nabla g(W) \\ &= \frac{1}{2} \sum_{i,j=1}^{2K} \Lambda_{K,ii}^{-1} D^{ij} g(W) (w'_i - w_i) (w'_j - w_j) + \frac{1}{4} \sum_{i,j,k=1}^{2K} \Lambda_{K,ii}^{-1} D^{ijk} g(W) (w'_i - w_i) (w'_j - w_j) (w'_k - w_k) \\ & \quad + \frac{1}{12} \sum_{i,j,k,l=1}^{2K} \Lambda_{K,ii}^{-1} D^{ijkl} g(t^*(W' - W) + W) (w'_i - w_i) (w'_j - w_j) (w'_k - w_k) (w'_l - w_l) \\ &= \frac{1}{2} \text{tr}[(W' - W)(W' - W)^\top \Lambda_K^{-\top} \nabla^2 g(W)] + \frac{1}{4} \sum_{i,j,k=1}^{2K} \Lambda_{K,ii}^{-1} D^{ijk} g(W) (w'_i - w_i) (w'_j - w_j) (w'_k - w_k) \\ & \quad + \frac{1}{12} \sum_{i,j,k,l=1}^{2K} \Lambda_{K,ii}^{-1} D^{ijkl} g(t^*(W' - W) + W) (w'_i - w_i) (w'_j - w_j) (w'_k - w_k) (w'_l - w_l), \end{aligned} \tag{3}$$

where $t^* \in [0, 1]$. Also by exchangeability,

$$\mathbb{E}[(W' - W)(W' - W)^\top] = 2 \mathbb{E}[W(W - W')^\top] = 2 \mathbb{E}[WW^\top \Lambda_K^\top] = 2\mathbf{V}\Lambda_K^\top.$$

It follows that

$$\mathbb{E}[\nabla^\top \mathbf{V} \nabla g(W)] = \mathbb{E} \text{tr}[\mathbf{V} \nabla^2 g(W)] = \frac{1}{2} \mathbb{E} \text{tr}[\mathbb{E}[(W' - W)(W' - W)^\top] \Lambda_K^{-\top} \nabla^2 g(W)]$$

Thus,

$$\begin{aligned} S &= \mathbb{E}[\nabla^\top \mathbf{V} \nabla g(W) - W^\top \nabla g(W)] \\ &= \frac{1}{2} \mathbb{E} \text{tr}[\mathbb{E}[(W' - W)(W' - W)^\top] \Lambda_K^{-\top} \nabla^2 g(W)] - \frac{1}{2} \mathbb{E} \text{tr}[(W' - W)(W' - W)^\top \Lambda_K^{-\top} \nabla^2 g(W)] \\ & \quad - \frac{1}{4} \mathbb{E} \sum_{i,j,k=1}^{2K} \Lambda_{K,ii}^{-1} D^{ijk} g(W) (w'_i - w_i) (w'_j - w_j) (w'_k - w_k) \\ & \quad - \frac{1}{12} \mathbb{E} \sum_{i,j,k,l=1}^{2K} \Lambda_{K,ii}^{-1} D^{ijkl} g(t^*(W' - W) + W) (w'_i - w_i) (w'_j - w_j) (w'_k - w_k) (w'_l - w_l). \end{aligned}$$

□

References

Chen, S. X., Zhang, L., and Zhong, P. (2010). Tests for high-dimensional covariance matrices. *Journal of the American Statistical Association*, 105(490):810–819.

- Cohn, D. L. (2013). *Measure Theory*. Birkhauser Advanced Texts Basler Lehrbucher. Birkhuser Basel, 2 edition.
- Goeman, J. J., van de Geer, S. A., and van Houwelingen, H. C. (2006). Testing against a high dimensional alternative. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(3):477–493.
- Hsu, D., Kakade, S., and Zhang, T. (2012). A tail inequality for quadratic forms of subgaussian random vectors. *Electronic Communications in Probability*, 17(0).
- Jiang, J. (1996). Reml estimation: asymptotic behavior and related topics. *Annals of Statistics*, 24(1):255–286.
- Kass, R. E. and Wasserman, L. (1995). A reference bayesian test for nested hypotheses and its relationship to the schwarz criterion. *Journal of the American Statistical Association*, 90(431):928–934.
- Maruyama, Y. and George, E. I. (2011). Fully bayes factors with a generalized g -prior. *Ann. Statist.*, 39(5):2740–2765.
- Shang, Z. and Clayton, M. K. (2011). Consistency of bayesian linear model selection with a growing number of parameters. *Journal of Statistical Planning and Inference*, 141(11):3463–3474.
- Vershynin, R. (2018). *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.